ARMY RESEARCH LABORATORY

Concept Evaluation of the Heterogeneous Data Proximity Tool at the C4ISR Network Modernization Event FY12

by Timothy P. Hanratty, Eric G. Heilman, John T. Richardson, and Mark R. Mittrick

ARL-TR-6783 **January 2014**

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5067

ARL-TR-6783 **January 2014**

Concept Evaluation of the Heterogeneous Data Proximity Tool at the C4ISR Network Modernization Event FY12

by Timothy P. Hanratty, Eric G. Heilman, John T. Richardson, and Mark R. Mittrick
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
January 2014	Final	March 2012–August 2012

4. TITLE AND SUBTITLE

Concept Evaluation of the Heterogeneous Data Proximity Tool at the C4ISR Network Modernization Event FY12

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Timothy P. Hanratty, Eric G. Heilman, John T. Richardson, and Mark R. Mittrick

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
ATTN: RDRL-CII-C
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-6783

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Visual analytic methods are a growing area of research that targets the effective interleaving of analytical reasoning with interactive interfaces. Capitalizing on the human capacity for spatial reasoning, visual analytics enhance the decisionmaker's understanding of supporting data by highlighting complex relationships. Toward this end, the U.S. Army Research Laboratory has developed a software application that complements traditional social network analysis, enabling improved understanding of the human terrain. The software application, called the Heterogeneous Data Proximity Tool (HDPT), combines a data dimensionality reduction routine using multidimensional scaling to produce a graphical representation of similarity or dissimilarity among a set of high-valued individuals. The HDPT starts with a reference data set representing individuals with known group affiliations: insurgents, innocents, and criminals. As intelligence is collected about new individuals within an area of operation, HDPT computes and plots their relative positions with regard to the existing reference data. The resulting analytic portrays the relative position of the new individual's organizational orientation within the known human terrain. This report provides an overview of the development of HDPT and presents the results from its participation in the Communications-Electronics Research, Development, and Engineering Center, Product Director Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance and Network Modernization's Event 2012 field exercise.

15. SUBJECT TERMS

visual analytic, multidimensional scaling, dimensional reduction, situation awareness, decision support

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Timothy Hanratty
Unclassified	Unclassified	Unclassified	UU	32	**19b. TELEPHONE NUMBER** (*Include area code*) 410-278-3084

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

The U.S. Army Research Laboratory (ARL) Tactical Information Fusion Branch (TIFB) focuses on the creation and application of data analysis techniques that assist Soldiers in improving military intelligence for command decisions. Challenging this effort is the unprecedented growth in technology that provides a large amount of data collected using a multiplicity of methods. Additionally, there is a limited understanding of human judgment and processes necessary to evaluate this data as an intelligence source, particularly in the presence of inconsistent, time-critical environments (*1*, *2*). To meet these challenges, TIFB is exploring visual analytics as an innovative method able to efficiently and effectively transform data into human-understandable information that imparts knowledge.

Visual analytics is a growing area of research that embeds analytical reasoning within interactive interfaces. Capitalizing on the human capacity for spatial reasoning, visual analytics enhance the decisionmaker's understanding of data by highlighting complex relationships (*3*, *4*). The application of visual analytics has yielded significant results in an array of paradigms including business, medicine, and defense (*1*, *5*). The TIFB has incorporated visual analytic techniques within a military decision software application called the Heterogeneous Data-reduction Proximity Tool (HDPT). HDPT provides a multidimensional analysis of nonobvious relationships among individuals through the calculation of their similarity compared with a known reference set. The HDPT visual analytic is a complement to traditional social network analysis able to improve a commander's understanding of human terrain by incorporating qualitative and quantitative information into the decisionmaking process.

This report documents a concept evaluation of the HDPT visual analytic application at the U.S. Army Research, Development and Engineering Command's Communications-Electronics Center Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) and Network Modernization Event 12 (E12). Section 2 of this report provides background on the statistical approach undertaken and the rationale for choosing the particular proximity calculation. In section 3, the system-level design and instantiation of the HDPT as a web application linked to the Distributed Common Ground Systems-Army (DCGS-A) is discussed. The implementation of HDPT in a U.S. Army field exercise and observations about its use are presented in section 4. The report concludes with lessons learned and the way forward in section 5.

2. Background

For many military applications, extracting knowledge from high-dimensional data sets is a persistent and complicated task. This is especially true when the data sets are of mixed data type, wherein the attributes defining the objects to be compared take on values from differing measurement scales. Moreover, the data of interest are typically amorphous; i.e., not linked to an explicit theory to assist the researcher in making inferences or predicting structure. To better incorporate these types of data into the decisionmaking process, HDPT uses multidimensional scaling (MDS) for visualizing data structure and Gower's similarity coefficient as the algorithm for calculating proximity matrices. The following sections provide a brief background on both MDS and Gower.

2.1 Multidimensional Scaling

Originating out of the fields of mathematical psychology and social sciences, MDS is a data analysis approach used to visually interrogate the similarity or dissimilarity between the pair-wise "distances" among a given set of objects (*6–11*). The values of the distances, sometimes called proximity measures or similarity measures**,** can be obtained either as perceived subjective measures or calculated objectively within the pair-wise comparison of the given set of objects. Most often, the objects are vectors of the form $X = (x_1, x_2, \ldots, x_m)$. The vector components x_k, collectively known as attributes, variables, or factors, provide the basis for comparison of objects. Given a similarity matrix for a set of objects, each object is projected as a point in n-space, arranged so the distances between the objects have the strongest possible relation to the similarity matrix. The intrinsic power of MDS is that it reduces complex n-space dimensionality, where n is the number of descriptive object attributes, to a human-interpretable two-dimensional (2-D) or 3-D space. It is the human-readable projection that promotes the exploratory analysis of data's hidden structure.

Differing from other forms of multivariate statistics, specifically principal component analysis, MDS does not constrain the data to be normally distributed. With that understanding, it becomes apparent that the hidden power behind meaningful MDS analysis is found in the construction of the similarity matrix and its projection into the reduced space. For calculating a similarity matrix, ideally all of the defining attributes should be of the same data type (*12*). Unfortunately, for many real-world problems, like the example in section 2.3, disparate scales of measure are commonplace, making matrix calculation problematic.

2.2 Gower Similarity Coefficient

One of the first to confront the combination of quantitative and qualitative (mixed scales of measure) was John Gower (*13*). Gower proffered that given an array of objects with *k* attributes, the global similarity value (S_{ij}) between two objects is defined as the summation of the individual attribute similarities (s_{ijk}) multiplied by a possible weighting factor. Here, s_{ijk} corresponds to the measure of local similarity assigned to the object pair (X_i, X_j) restricted to attribute *k*. The summation of the individual similarities is divided by the summation across all weights. Gower's similarity coefficient equation, equation 1, allows for the weighing of individual attributes and the possibility of missing data.

$$S_{ij} = \frac{\sum_{1}^{k} s_{ijk}\, w_k}{\sum_{1}^{k} w_k}\text{; } w_k \text{ weight assigned to an individual attribute} \tag{1}$$

The classic calculation for individual similarities is shown in equation 2, where X_{ik} and X_{jk} are the *k*th attribute for objects X_i and X_j, respectively. R_k is defined as the range for that particular quantitative attribute. In recent years, numerous extensions to similarity measurement calculations have been attempted in a wide array of subject areas, from image processing to medical informatics. Approaches taken include, but are not limited to, rough sets (*14*), fuzzy logic (*15*), and ordinal extensions (*16*).

$$s_{ijk} = \begin{cases} \begin{cases} 1, & \text{if } x_{ik} = x_{jk} \\ 0, & \text{if } x_{ik} \neq x_{jk} \end{cases} & \text{, k is qualitative} \\ 1 - \dfrac{|x_{ik} - x_{jk}|}{R_k} & \text{, k is quantitative} \end{cases} \tag{2}$$

2.3 Illustrative Use Case

In this section, the use of Gower's similarity coefficients and their application with an MDS interface are illustrated with a sample data set of terrorist activity recorded in table 1. Eleven sample terrorist events are characterized by the five features (attributes) listed in the first row and their respective scale of measurement in the second row.

Table 1. A representative set of terrorist event data (*17*).

Attribute	**Day**	**Location**	**Time**	**Primary Attack**	**Secondary Attack**
Scale:	Nominal	Nominal	Interval	Nominal	Binary
Event 1	Saturday	Alpha sector	1800	SAF	No
Event 2	Wednesday	Charlie sector	1200	IED	Yes
Event 3	Saturday	Alpha sector	1900	SAF	No
Event 4	Saturday	Bravo sector	1500	VBIED	No
Event 5	Wednesday	Charlie sector	0600	IED	Yes
Event 6	Saturday	Bravo sector	1800	SAF	No
Event 7	Wednesday	Charlie sector	1100	VBIED	Yes
Event 8	Tuesday	Echo sector	1900	VBIED	No
Event 9	Wednesday	Delta sector	1100	IED	Yes
Event 10	Thursday	Foxtrot sector	1000	VBIED	Yes
Event 11	Sunday	Delta sector	2000	VBIED	Yes

Notes: SAF = small arms fire; IED = improvised explosive device; VBIED = vehicle-borne IED.

Each event is defined by the day of the week it occurred, the location and time of the event, the type of primary attack the event employed, and whether a secondary attacked occurred at the same time. In this simple example, the day of the week characteristic is not considered chronologically, and therefore an ordinal relationship was not used. In point of fact, both the day of the week and location are treated as nominal data. The times of attack are recorded on a 24-h clock, providing an interval/ratio scale of measure. The primary attack modes are as follows: small arms fire (SAF), improvised explosive device (IED), and vehicle-borne IED (VBIED). These modes are multilevel nominal. Threats will sometimes initiate a secondary attack during an event, and this is recorded as binary data type.

The assessment of the similarity between event 1 and event 2 using Gower's general coefficient requires the evaluation of $S_{12} = {}_{k=1}\sum^{5} w_{12k}\, s_{12k} / {}_{k=1}\sum^{5} w_{12k}$. A description follows of the local similarities, s_{ijk}, between the event pair (i, j) for attributes *k*, as defined by equation 2.

For events 1 and 2, the local similarities s_{12k} , $k = 1, \ldots, 5$, take on the following values.

$s_{121} = 0$ Saturday $\neq$ Wednesday

$s_{122} = 0$ Alpha sector $\neq$ Charlie sector

s_{123} = 0.75   1 − |18 − 12| /24 = 0.75

s_{124} = 0   SAF $\neq$ IED

s_{125} = 0   no $\neq$ yes

The global similarity between events 1 and 2 is then calculated as

$$S_{12} = (0 + 0 + 0.75 + 0 + 0) / 5 = 0.15 \tag{3}$$

with corresponding dissimilarity $\sqrt{(1 - S_{12})}$ = 0.922 (*17*).

Because Gower's coefficient is a similarity score, $1 – S_{ij}$ is the corresponding dissimilarity. Appropriately normalized, both take on complementary values in the unit interval [0, 1]; that is, similarity + dissimilarity ≡ unity. The mapping $\sqrt{(1 - S_{12})}$ serves simply to enlarge the small values that will always be encountered because they can never exceed unity, and it has no impact on the final result.

The application of Gower's algorithm to all of the terrorist events is shown in table 2. These values represent the dissimilarity measures between the events found in table 1 and the precursors to exercising the multidimensional scaling algorithm.

Table 2. Dissimilarity coefficients for the events shown in table 1.

0	.9219	.4564	.6519	.8366	.4472	.9264	.7799	.9246	.9309	.9036
.9219	0	.8113	.9082	.5000	.9219	.4564	.9264	.4564	.7852	.8164
.4564	.8113	0	.6582	.7011	.6390	.9309	.7745	.8164	.9354	.8990
.5619	.9082	.6582	0	.8215	.4743	.7958	.6582	.9128	.8010	.8010
.8366	.5000	.7011	.8215	0	.8366	.6845	.8316	.6645	.9128	.9398
.4472	.9219	.6390	.4743	.8366	0	.9264	.7799	.9264	.9309	.9036
.9264	.4564	.9309	.7958	.6645	.9264	0	.8164	.6324	.6390	.6892
.7799	.9264	.7745	.6582	.8316	.7799	.8164	0	.9309	.8215	.7799
.9264	.4564	.8164	.9128	.6645	.9264	.6324	.9309	0	.7799	.6892
.9309	.7852	.9354	.8010	.9128	.9309	.6390	.8215	.7799	0	.6952
.9036	.8164	.8990	.8010	.9398	.9036	.6892	.7799	.6892	.6952	0

Given the calculated dissimilarity matrix for a set of terrorist events, each event is projected as a point in n-space, arranged so the distances between the events have the strongest possible relation to the dissimilarity matrix. Figure 1 represents the visual analytic for the terrorist event data set.

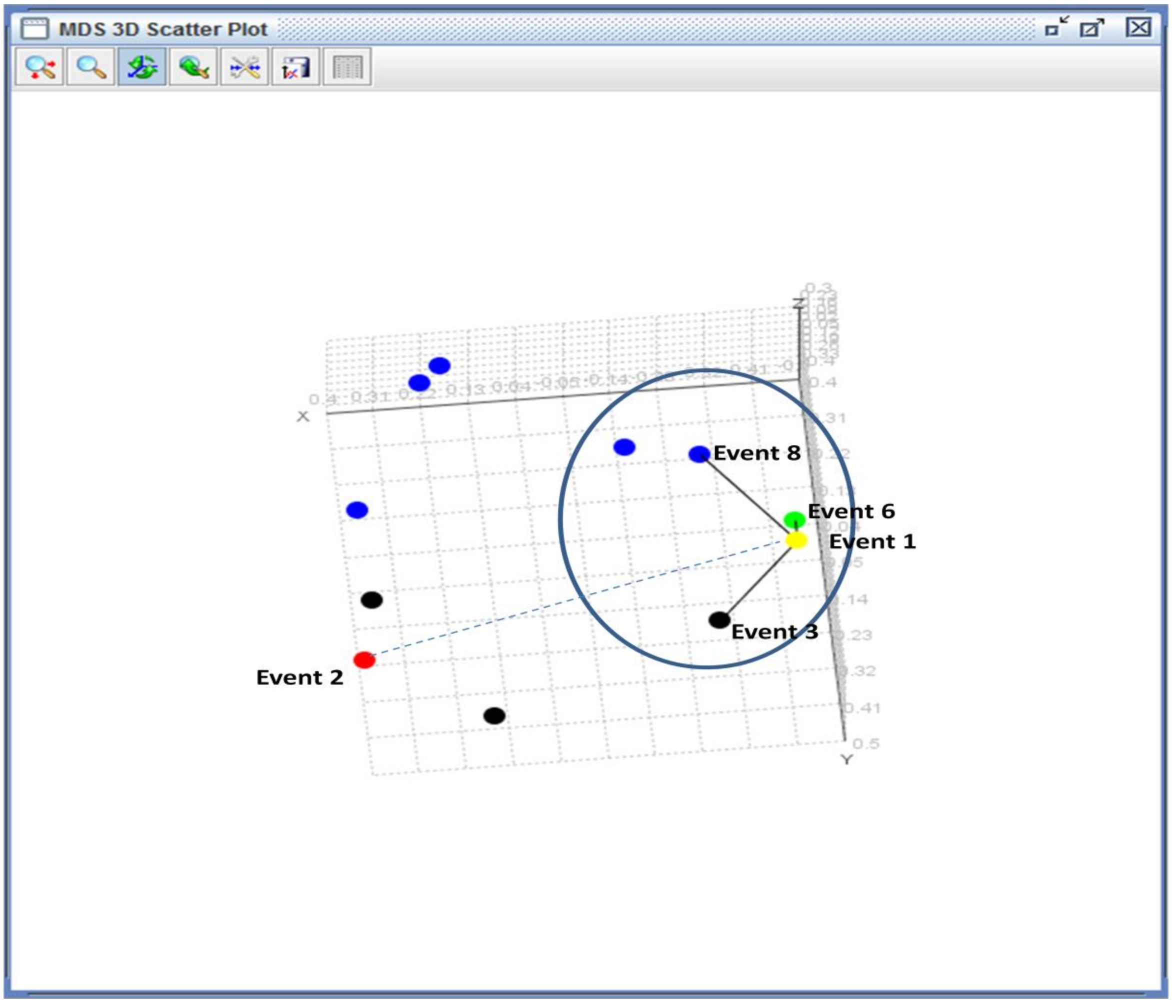


Figure 1. 3-D representations of the event data from table 1.

Reducing high-dimensional data sets into a lesser-dimensional visual analytic assists the human in comprehending larger data sets and provides the opportunity for gaining valuable insights and exploitation. In this example, each node represents one of the 11 terrorist events listed in table 1. The distances between and among the nodes mimic the elements of the dissimilarity matrix. For example, it appears in figure 1 that event 1 differs widely from event 2 with the corresponding dissimilarity matrix entries—0.9219 confirming this observation. The visualization also reveals an interesting clustering of events 1, 3, 6, 8, and 4, where each of these events had an associated secondary attack as one of its characteristics. Additionally, a closer inspection of event 1 reveals that its three most similar events (3, 6, and 8) share the characteristics of occurring around 1800 and were initiated as SAF events. Interpretation of results is a crucial part of the process.

The following section details the design and development of the HDPT visual analytic application; specifically, the HDPT's development for concept demonstration in the U.S. Army's C4ISR On-The-Move (OTM) 2012 exercise (E12) and associated scenario.

3. HDPT System Overview

While there have been research efforts into the development of qualitative and quantitative similarity analysis, few have been effectively coupled with a visualization framework, and fewer still have been interactively coupled with a tactical military decision support tool. Toward that end, HDPT is the software instantiation of a visual analytic technique that effectively combines MDS with the flexibility of the mixed-scale Gower similarity calculation. The targeted area of interest for this instantiation is the assessment of individuals within a tactical social network. This specific implementation was designed as a concept demonstration for the Product Director C4ISR and Network Modernization's E12 exercise and was integrated with the DCGS-A program as data feed.

At a high level of abstraction, the concept of operation for this exercise was as follows. The HDPT started with a reference data set that represented the characteristics of individuals with known group affiliations: insurgents, innocents, and criminals. As intelligence data was collected about new individuals within an area of operation, HDPT computed the individuals' similarity to the reference data set and plotted their relative positions in an associated 3-D visualization space. The resulting analytic portrayed the relative position of the new individual's orientation within the known human terrain (insurgent versus innocent versus criminal). The goal of the tool is to assist a military analyst with an improved understanding of the local human environment and in defining future information requests.

Shown in figure 2 are three major components that make up the HDPT system:

- HDPT Web Application
- DCGS-A Global Graph
- statistics engine

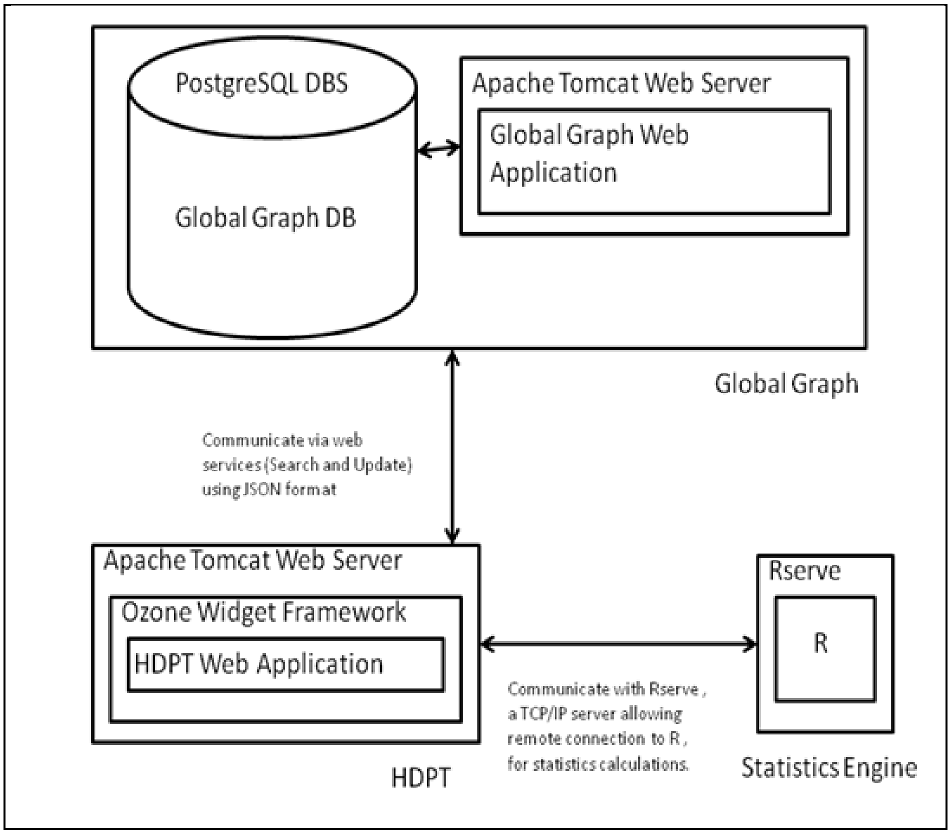


Figure 2. HDPT system diagram.

HDPT is a Web application that is deployable within the Ozone Widget Framework (OWF). HDPT accesses its data source, the DCGS-A Global Graph, via a Web service. To calculate similarity, it uses the Gower Similarity and Multidimensional Scaling algorithms contained in the "R" statistical computing environment. The Global Graph and OWF are both products of Potomac Fusion, Inc. and part of the DCGS-A program of record. The principal HDPT component and subject of the remainder of this paper is the HDPT Web Application. The following subsections outline the design of the HDPT Web Application as demonstrated at the C4ISR E12 exercise.

3.1 HDPT Web Application

The HDPT Web Application is the core of the HDPT system and serves as the user interface for performing similarity analysis. The primary components of the HDPT Web Application used in E12 are displayed in figure 3 and include the following: (1) HDPT Menu Bar, (2) Search Window Panel, and (3) Plot Window Panel.

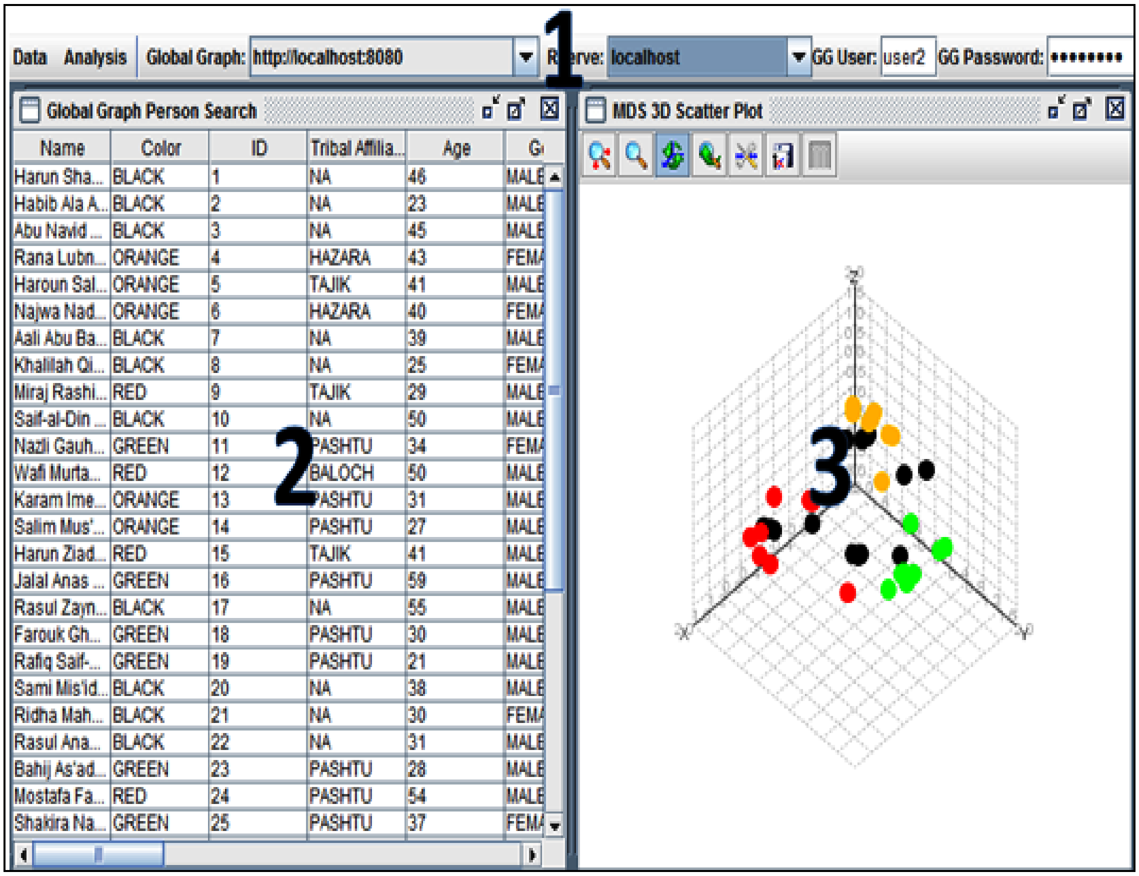


Figure 3. HDPT web application components: (1) Menu Bar, (2) Search Window, and (3) Plot Window.

3.1.1 HDPT Menu Bar

The HDPT Menu Bar, shown in figure 3, contains drop-down selections for loading data, creating the visual analytic, and configuring the HDPT. The Menu Bar allows access to the Data Menu, Analysis Menu, and the configuration components.

- The Data Menu contains drop-down selections for loading data into the tool and propagating updates back to the data source. The data source used by HDPT during the E12 exercise was the DCGS-A Global Graph. In this exercise, a structured query language (SQL) version of the Global Graph was used that consisted of a PostgreSQL database and associated Web services for searching and updating the database. A Representational State Transfer (REST) Web service protocol was used for communication between HDPT and the data source via a JavaScript Object Notation (JSON) data structure. REST is a lightweight alternative protocol to mechanisms like the Simple Object Access Protocol, remote procedure call, or Constraint-Based Reconstruction and Analysis. Typically with REST, Hypertext Transfer Protocol is used to make the connections. Likewise, JSON is a lightweight data-interchange format designed for the transport of structured text. The election of these protocols greatly facilitated connection and interaction with the DCGS-A framework.

- The Analysis Menu contains drop-down selections for plotting the visual analytic as an interactive 3-D scatter plot and setting user preferences. The Plot Menu selection becomes available once there is an active search window panel. In addition to plotting, the Analysis Menu allows two user preferences to be modified: links displayed and attribute threshold. Both of these preferences change the way links between nodes in the plot are handled. By default, when interacting with the plot, the user can right-click on a node and links will be drawn to the three most similar reference nodes; using the "links displayed" preference, the number of links drawn can be increased or decreased. Additionally, entities with four or more attribute values are displayed as nodes within the visual analytic; using the "attribute threshold" preference, the minimum number of attribute values necessary for node plotting can be adjusted.
- The Configuration Menu allows the user to specify server parameters for the Global Graph and R server. There is a drop-down menu for each server selection that contains the internet addresses of several commonly used server host machines for the OTM exercise. In addition, mandatory fields are provided for users to enter their username and password for Global Graph access. HDPT uses these values for connecting to both the Global Graph REST Web services and the R statistics engine.

3.1.2 Search Window Panel

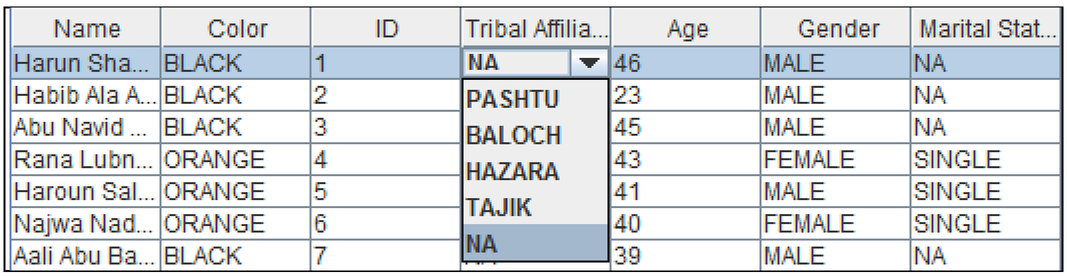
The HDPT search window panel displays the dataset returned from a Global Graph search. Figure 4 shows an example of this search window.

Global Graph Person Search

Name	Color ▲	ID	Tribal Affiliat...	Age	Gender	Marital Status
Aali Abu Bakr Karim	BLACK	11	NA	39	MALE	NA
Abu Navid Sultan	BLACK	32	PASHTU	42	MALE	NA
Amirah Sani El-Amin	BLACK	36	NA	48	FEMALE	NA
Habib Ala Ahmed	BLACK	34	NA	23	MALE	NA
Harun Shahzad El-Mofty	BLACK	9	NA	46	MALE	NA
Hussain Mansoor El-Hashem	BLACK	22	NA	32	MALE	NA
Ikram I'timad Abdullah	BLACK	16	HAZARA	20	MALE	MARRIED
Khalilah Qismat Amirmoez	BLACK	13	NA	25	FEMALE	NA
Rasul Anass Zaman	BLACK	4	NA	31	MALE	NA
Rasul Zayn Mohammed	BLACK	20	NA	55	MALE	NA
Ridha Mahdi El-Mofty	BLACK	19	NA	30	FEMALE	NA
Saif-al-Din Jinan Hakim	BLACK	24	NA	50	MALE	NA
Sami Mis'id El-Ghazzawy	BLACK	5	NA	38	MALE	NA
Yusuf Mehmud Samara	BLACK	15	NA	42	MALE	NA
Zaman Noor Hakim	BLACK	7	NA	34	MALE	NA
Bahij As'ad Tawfeek	GREEN	6	PASHTU	28	MALE	MARRIED
Farouk Ghayth El-Ghazzawy	GREEN	26	PASHTU	30	MALE	MARRIED
Gabr Hussein Ahmed	GREEN	1	BALOCH	35	MALE	SINGLE
Hashim Fouad Admad	GREEN	3	TAJIK	35	MALE	SINGLE
Jalal Anas Kader	GREEN	18	PASHTU	59	MALE	SINGLE
Nazli Gauhar Ajam	GREEN	8	PASHTU	34	FEMALE	MARRIED
Rafiq Saif-al-Din Karim	GREEN	33	PASHTU	21	MALE	SINGLE
Shakira Nashwa Abujamal	GREEN	37	PASHTU	37	FEMALE	MARRIED
Haroun Salih Abdullah	ORANGE	39	TAJIK	41	MALE	SINGLE
Jinan Qadir El-Ghazzawy	ORANGE	27	BALOCH	44	MALE	SINGLE
Karam Imen Boulos	ORANGE	29	PASHTU	31	MALE	SINGLE
Majid Rusul Abujamal	ORANGE	25	TAJIK	38	MALE	SINGLE
Mostafa Tufayl Karimi	ORANGE	28	PASHTU	30	MALE	SINGLE
Najwa Nadia Saqqaf	ORANGE	17	HAZARA	40	FEMALE	SINGLE
Rana Lubna Samara	ORANGE	12	HAZARA	43	FEMALE	SINGLE
Salim Mus'ad Hakim	ORANGE	2	PASHTU	27	MALE	SINGLE
Adam Abdur-Rashid Abdullah	RED	10	BALOCH	35	MALE	MARRIED
Harun Ziad Boulos	RED	14	TAJIK	41	MALE	MARRIED
Miraj Rashid Karimi	RED	21	TAJIK	29	MALE	MARRIED
Mostafa Farooq Darzi	RED	38	PASHTU	54	MALE	MARRIED
Nasir Baki Saab	RED	23	HAZARA	34	MALE	SINGLE
Sulaiman Badr Muhammad	RED	31	PASHTU	44	FEMALE	SINGLE
Wafi Murtada Hakim	RED	30	BALOCH	50	MALE	MARRIED
Ziyad Guda Sultan	RED	35	HAZARA	34	MALE	MARRIED

Figure 4. Search window panel.

The window consists of a table where each row is a person and each column is an attribute. The result returned from the Global Graph Web search service is a JSON structure containing matching people. HDPT processes the JSON structure into a separate Java object for each person. The search window has a number of user interactive features. The attribute editing feature allows a user to change the values of an entity's attributes. For the E12 exercise, users were able to change attributes using precoded drop-down menus available in each cell of the search panel window, as shown in figure 5. The search window panel allows rows to be sorted according to the values in any column. Clicking the mouse on the column heading will cause the rows to sort alphabetically (words) or number order (digits) according to the data in that column. For example, the search panel shown in figure 4 is sorted by the Color column. Finally, if the plot window is currently active, clicking on a row in the search window will highlight the node that corresponds to that row in the 3-D scatter plot; changing that node's color to a neutral yellow.



Name	Color	ID	Tribal Affilia...	Age	Gender	Marital Stat...
Harun Sha...	BLACK	1	NA	46	MALE	NA
Habib Ala A...	BLACK	2		23	MALE	NA
Abu Navid ...	BLACK	3		45	MALE	NA
Rana Lubn...	ORANGE	4		43	FEMALE	SINGLE
Haroun Sal...	ORANGE	5		41	MALE	SINGLE
Najwa Nad...	ORANGE	6		40	FEMALE	SINGLE
Aali Abu Ba...	BLACK	7		39	MALE	NA

Figure 5. Example of editing attributes within the search window.

3.1.3 Plot Window Panel

The plot window panel displays a 3-D scatter plot of MDS results. The development of the visualization used the JMathPlot open-source graphics library. A sample plot window panel from the E12 exercise is shown in figure 6. In this case, the node under investigation is highlighted using a neutral yellow color and has the three most similar reference nodes linked with straight lines, sharing similarity between two nodes from the criminal set (orange) and one from the friendly set (green).

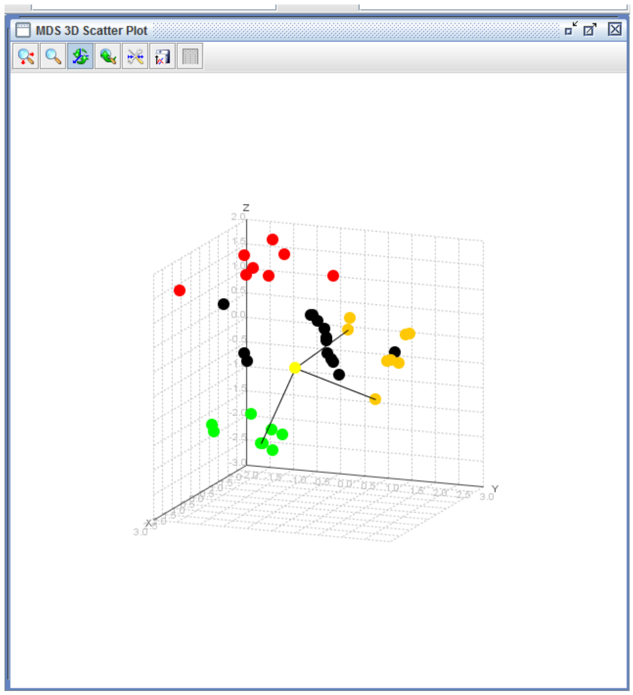


Figure 6. Plot window panel showing similarity links.

3.2 HDPT Plot Window Panel Utility

In the E12 exercise, the nodes were color coded with the following population schema: black = unknown, red = insurgent, green = friendly, and orange = criminal. As the exercise transpired, data describing unknown persons were updated in the Global Graph, either from the HDPT console or a mobile device ARL developed for Soldiers' use during field encounters. As the data associated with each black node changed, the HDPT visual analytic also changed. Soldiers participating in the exercise as intelligence analysts were tasked with using the scatter plot to determine the disposition of black nodes. That is, did projection of the dimensionally reduced characteristics of each black node more likely belong to the red, green, or orange population as determined by proximity (similarity)?

To assist the Soldiers' understanding of the underlying decision space, HDPT provided a couple of important capabilities within the 3-D visual analytic that were used extensively throughout the E12 exercise. First, as shown in figures 7 and 8, HDPT provided the ability to freely rotate the decision space along any axis. Projection of a 3-D decision space onto a 2-D screen can be problematic; objects that appear close to one another in 2-D can actually be far apart. The ability to rotate along any axis was critical to correctly interpreting the relation projections of the decision space.

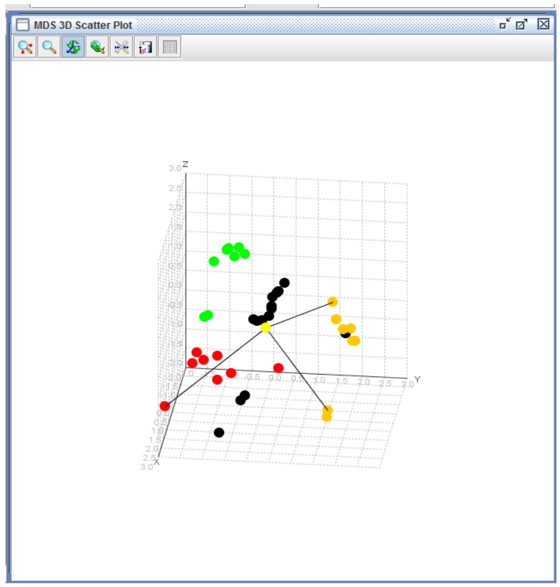


Figure 7. HDPT view before rotation.

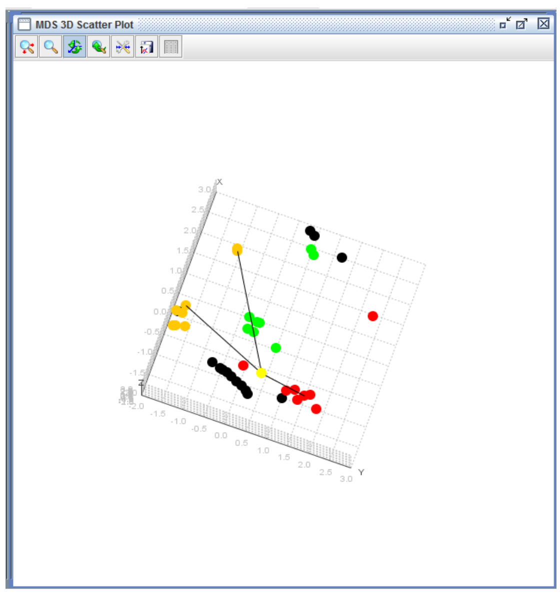


Figure 8. HDPT rotated view.

A second capability that was used at length was the zoom. As shown in figures 9 and 10, the zoom capability permitted users the ability to examine in finer detail the related nodes that were clustering close to the node in question.

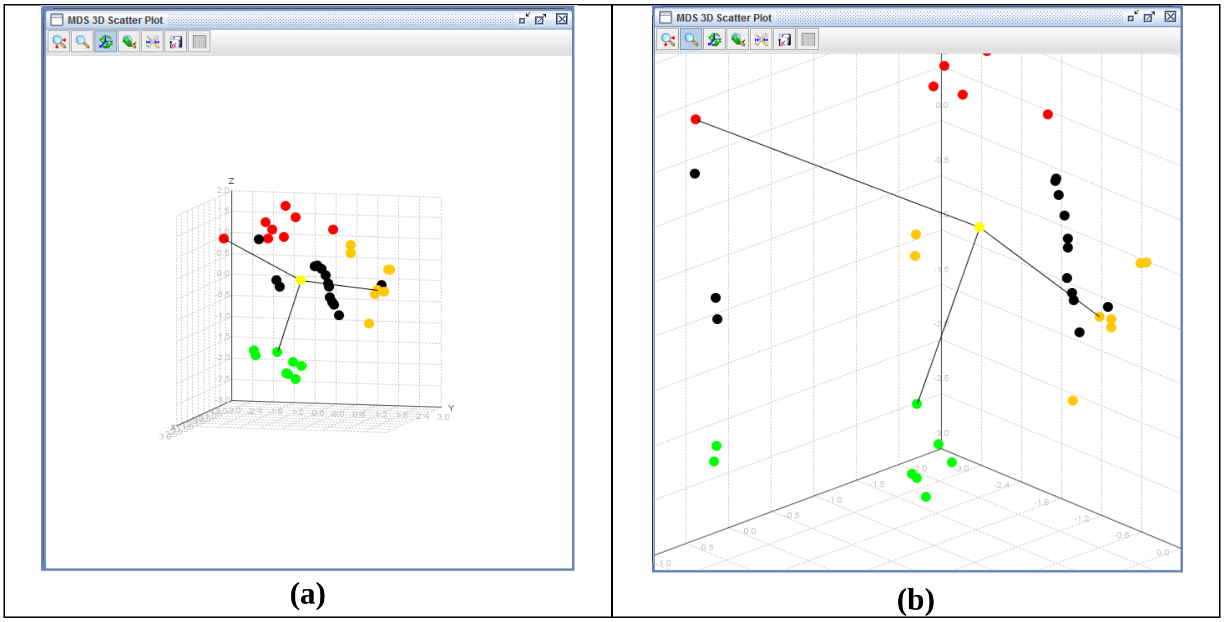


Figure 9. HDPT view before zoom (a) and HDPT after zoom (b).

Name	SubID	Location	Org	Age	Gen	TA	MS	Nat	POB	Equip	VehicleID	CR	Ed	Emp	MR	Rel	Sk	Add
Bahij As'ad Tawfeek		39.98 / -74.43	Friendly	28	M	Pastun	M	HN	BIA	Uniform	Burgundy Luxury Sedan	No	High	WC	SS	Mld	ME	TSV

Figure 10. Example of a person data element set (*18*).

In this example, the node in question (yellow) shows a high similarity to three neighboring nodes—one from each of the categorical types.

The plot window panel also offered several interactive features to assist in exploratory analysis. Left-clicking on a node in the display both highlighted it and the corresponding row in the search window so that the node is referenced back to its original data. Second, as discussed previously, right-clicking on a node will draw links from itself to three or more of its most similar reference node neighbors.

4. HDPT C4ISR E12 Concept Evaluation

The goal of the C4ISR E12 event was to provide stakeholders from across the Department of Defense to assess next-generation technologies. The annual event, held at Fort Dix, NJ, offers researchers a military relevant venue to assess, evaluate, and validate emerging technologies and facilitate technology maturation and transition to the acquisition process. The following section details HDPT participation at the E12 event.

4.1 Reference Data Set Description

The HDPT team created a multiple-dimensioned attribute data set describing persons affiliated with several groups presumed to be operational in the E12 exercise area. The groups included personnel considered to be nonhostile or friendly, overtly hostile (insurgent terrorists), and locally hostile (criminal). The attribute set, shown in table 3, represents information collected through combat questioning of individuals met during field encounters. To support HDPT analysis, each attribute can take the form of one or more specific values. For instance, the Martial Status attribute has a value of either Married or Not Married to represent the current condition of an individual. A full set of specific values for a person's attributes constitutes a unique data set representing that specific individual. An illustrative example of a person data element set is shown in figure 9. The subject's name is Bahij As'ad Tawfeek, a male subject. He is 28 years old, married, and a member of the Pastun tribe. Bahji was born in the area and is an Afghan national (HN is Host Nation). Soldiers have observed him wearing a military-like uniform and associated him with a burgundy luxury sedan. He has no criminal record, is highly educated, and is considered to be holding down a white collar job as a mechanic serving in the local military. He belongs to a milder religious sect where he lives in Times Square Village (*18*).

Table 3. HDPT data attributes.

Tribal Affiliation	Education Level
Marital Status	Employment Type
Nationality	Military Record
Place of Birth	(person religion) Member of
EquipmentID	Skill
VehicleID	AddressID
Criminal Record	—

To test the HDPT, the ARL team created 39 attribute sets, similar to figure 10. These were broken down into three reference groups, each containing eight individuals representing friendly, insurgent, and criminal organizations, and 15 sets representing initially unknown individuals. The reference groups formed node clusters in the HDPT visual plot, with each node representing one person. Figure 11 shows the HDPT display with both the insurgent (red nodes) and the friendly (green nodes) organizations circled. The black nodes represent unknown individuals. For ground-truth purposes, the full set of attributes for each of the 15 unknown persons placed their representative black nodes within an HDPT reference cluster. Removing data from the 13 fields listed in table 2 causes the neutral alignment of black nodes for the initial HDPT display shown in figure 12.

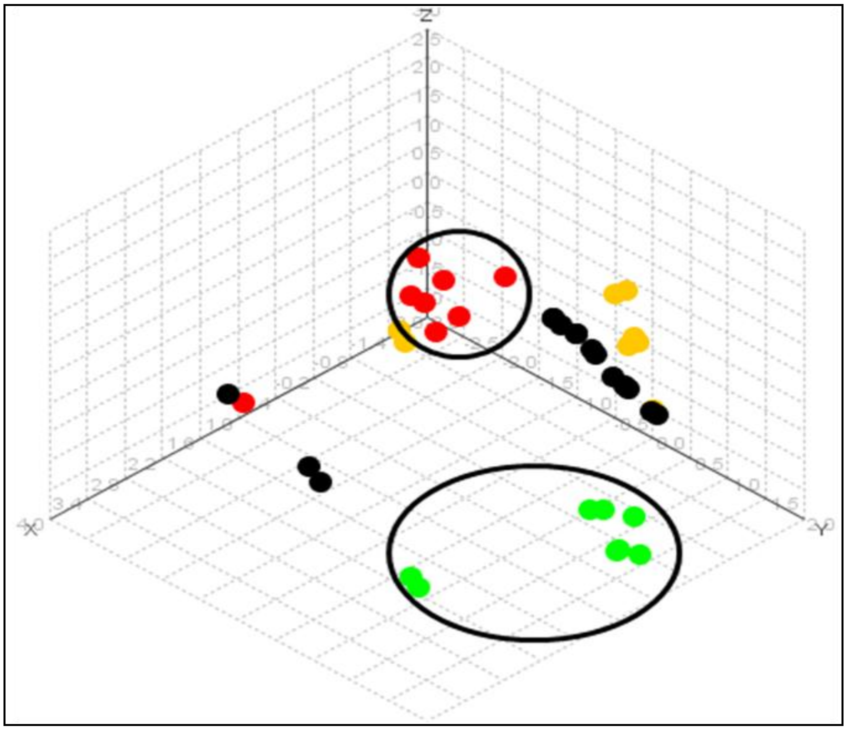


Figure 11. HDPT reference group visual display.

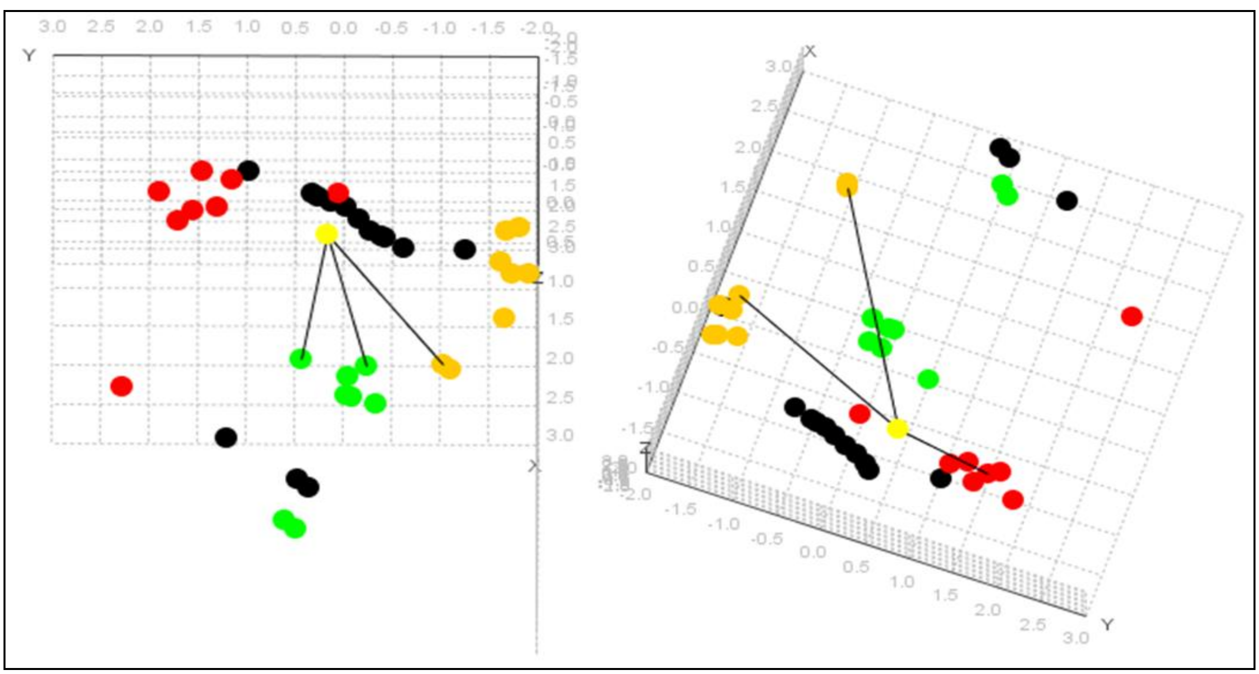


Figure 12. Progressive data discovery for a single unaligned person.

4.2 Scenario Inputs

During each day of the field exercise, Soldiers conducted a checkpoint and a presence patrol. During these missions, up to four threat actors (unknown individuals) were subjected to combat questioning, which was used to obtain data to populate the character's 13 missing fields. Early in the exercise, the ARL team found that the realistic conditions of combat questioning most often yielded only a part of the data necessary for full analysis. To support a thorough study of HDPT capabilities, the ARL team injected data for eight unused characters to ensure that an optimal set of attributes was available for analysis by Soldiers performing the military intelligence functions within the tactical operations center. Four of the characters were represented using the complete set of 13 data attributes while the other four characters had data for only 6 attributes. The resulting data mix provided an ad hoc condition from the field data as well as a controlled condition supportive of HDPT ground-truth analysis.

Data injects were divided into two sections, one each for the morning afternoon missions. Figure 12 shows an example visual analytic of a daily progression of data found for a single character. The character node is colored yellow and has the three "most similar" nodes linked with straight lines. The left visual analytic represents data discovered in the morning of an exercise day while the visual analytic on the right is an accumulation of the morning data and data discovered during the afternoon. The progression of data discovery is representative of intelligence accumulation on persons of interest over a period of time. In this example, the morning data causes the character to appear friendly, but additional data from afternoon collection casts the character more as a criminal. The ground truth puts this character in the criminal organizational group.

4.3 Soldier Survey

To track the progression of the Soldiers' understanding throughout an exercise day, a survey form (see appendix) was used. Using the HDPT visual analytic, each Soldier was asked to give an initial, mid-morning, and mid-afternoon assignment for each of the unknown (black nodes) individuals' association with a group and to weight that assignment on a scale from 1 to 5. The Soldiers' assessment as compared with character ground truth was used as a measure of HDPT's utility. Soldiers were also given the chance to comment on their understanding of situational awareness based on the HDPT visual analytic, the utility of the tool, and open comments for improvements.

4.4 Results

During the exercise, there were 11 different Soldiers acting as intelligence analysts using the HDPT. The raw survey scores are shown in table 4. Ten of the 11 Soldiers were infantrymen, typical of those found on a Company Intelligence Support Team. Only 2 of the 11 had military intelligence training, while most felt confident in their computer use abilities. Using a Likert Scale of 1–5 (where 1 = very poor and 5 = very well), the survey of the Soldiers revealed the following data concerning the use of the HDPT, the effectiveness of HDPT to track the evolving High Value Individual intelligence picture, and their opinion of the usefulness of HDPT in a tactical deployment. Most found that HDPT was easy to use, scoring usability 4.18 of a possible 5. The Soldiers also thought that HDPT addresses a tactically useful function, scoring a 4.09 of a possible 5. Some Soldiers were concerned about HDPT's ability to provide an improved understanding of civilian personnel within the exercise area of operations, scoring a 3.8 of a possible 5. Concerns raised during the event were reflected in the survey's comments section.

Table 4. HDPT survey result table.

Soldier No.	Age	MOS	Rank	Position	Years In	Color Blind?	Comptuer Use	No. of Deployments	Intel. Training	How Well Used	Unknown Person Sit. Aware	HDPT Use
1	24	19K30	SGT	Tank cdr.	5	No	2 of 5	2	None	4 of 5	4 of 5	4 of 5
2	21	11B	PV2	SAW	0.75	No	5 of 5	0	None	5 of 5	3 of 5	3 of 5
3	26	11B	SGT	Squad ldr.	5	No	4 of 5	2	None	4 of 5	3 of 5	3 of 5
4	24	11B	SPC	Team ldr.	5	No	4 of 5	1	None	4 of 5	4 of 5	3 of 5
5	19	11B	PFC	Infantryman	1	No	3 of 5	0	None	4 of 5	4 of 5	5 of 5
6	41	11B	SGT	Squad ldr.	18.5	No	5 of 5	7	Yes	4 of 5	5 of 5	4 of 5
7	52	42A	LTC	S3	32	No	3 of 5	2	Yes	5 of 5	—	5 of 5
8	21	11B	PV2	Infantryman	1	No	4 of 5	0	None	5 of 5	4 of 5	4 of 5
9	22	11B	PFC	Infantryman	1	No	4 of 5	0	None	3 of 5	4 of 5	4 of 5
10	19	11B	PFC	Infantryman	1	No	4 of 5	0	None	4 of 5	4 of 5	5 of 5
11	20	11B	PV2	SAW	0.75	No	3 of 5	0	None	4 of 5	3 of 5	5 of 5
—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	Averages =		3.72	—	—	4.18	3.8	4.09

Notes: MOS = military occupational skill; SAW = squad automatic weapon; S3 = operations;

When commenting on the HDPT, Soldiers suggested the development of several features that were implemented during the exercise, including (1) connection of the three persons closest to an unknown node under examination with straight line upon querying and (2) an improved resolution of nodes that overlap in the HDPT visual analytic. Additionally, Soldiers suggested several improvements that are being developed currently, to include (1) highlighting of the HDPT person data search window panel corresponding to both a chosen node and (2) the three closest nodes to that node and entity node rotation while in a zoomed portion of the visual analytic. As these suggestions originate from the ultimate field user of the HDPT, each will strengthen the relevance of the final product. The value of such interactions early within the technology creation process is crucial to ensure the creation of the highest quality tools for our Soldiers' use.

Soldiers using the HDPT examined the similarity of each unknown node with two goals: (1) predict in which reference group an unknown node belonged and (2) provide a confidence level for that prediction from 1 (least confident) to 5 (most confident). There were two different

Soldiers using HDPT during each day of the exercise. As shown in figure 13, the Soldiers' analysis of nodes with both full and partial attribute data resulted in a high percentage of correct predictions. In fact, most analysis became progressively more correct or remained at a high level of correctness throughout an exercise day. During the last four days of the exercise, Solders using the HDPT visual analytic correctly predicted a node's ground truth with an overall 93% accuracy.

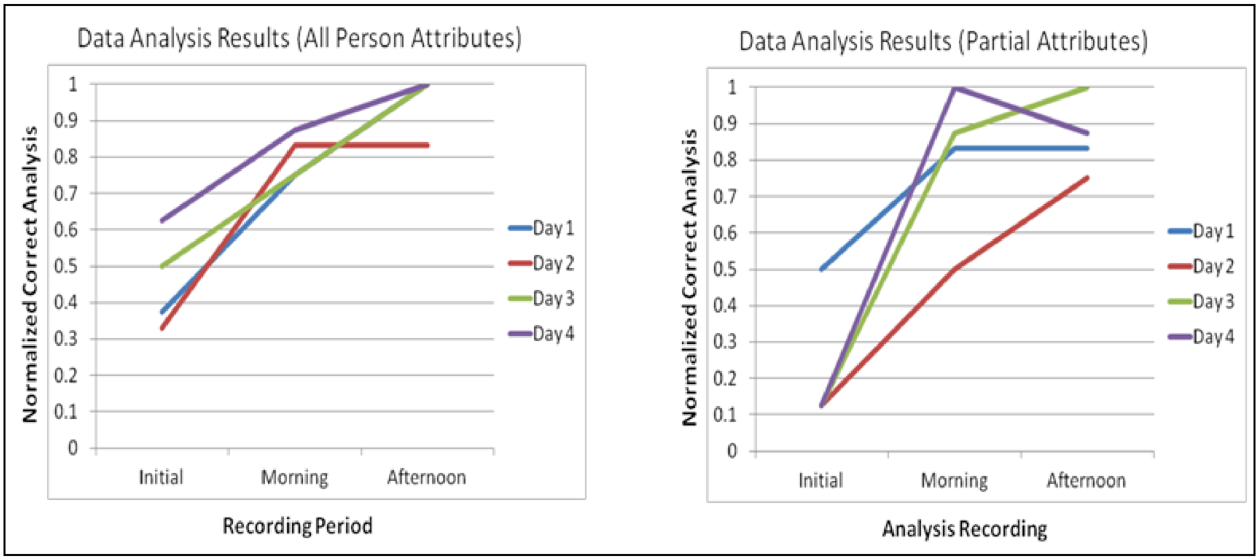


Figure 13. Soldier predictive analysis using the HDPT visual analytic compared with ground truth.

5. Conclusion

MDS is a powerful algorithm with the potential to enhance military intelligence analysis techniques such as social network analysis, cluster analysis, and pattern recognition (*19*). To that end, HDPT was developed as a utility for understanding human terrain by incorporating MDS similarity analysis rendered in a visual analytic display. The development of HDPT using tactical DCGS-A capabilities, namely the Ozone Widget and Global Graph environments, made it possible to demonstrate this technology at the E12 exercise.

TIFB developed an E12 fielding plan and scenario meant to rigorously test the power of similarity analysis and to solicit Soldier inputs for continued HDPT development. Soldier inputs enabled the E12 team to develop and implement new capabilities even as the exercise was occurring. Results of the testing indicate that HDPT has the potential, under the right conditions, to aid the determination of personnel group affiliation when encountered during field events.

The insights gained during E12 have led to important extensions of this work. Specifically, the development of formal procedures to determine the value of information collected in context of the operational tempo and the information's content and source reliability (*20*). Additionally, TIFB will use information gained to improve both the HDPT user interface and visual analytic. Building on E12 successes, the newly enhanced HDPT will be tested with Soldiers trained as military intelligence analysts during the upcoming E13 field exercise.

6. References

1. Thomas, J. J.; Cook, K., Eds. *Illuminating the Path: The R&D Agenda for Visual Analytics*. IEEE Computer Society Press: Washington, DC, 2005, p 4.

2. Greitzer, F. L.; Noonan, C. F.; Franklin, L. R. *Cognitive Foundations for Visual Analytics*. Pacific Northwest National Laboratory (PNNL): Richland, WA, 2011.

3. Börner, K.; Chen, C.; Boyack, K. Visualizing Knowledge Domains. In *Annual Review of Information Science & Technology*, Cronin, B., Ed., Information Today, Inc./American Society for Information Science and Technology: Medford, NJ, 2003; vol. 37, ch. 5, pp 179–255.

4. Heer, J.; Card, S.; Landay, J. Prefuse: A Toolkit For Interactive Information Visualization. Presented at the SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, 2–7 April 2005.

5. Hanratty, T.; Hammell II, R.; Yen, J.; McNeese, M.; Oh, S.; Kim, S.; Minotra, D.; Strater, L.; Cuevas, H.; Colombo, D. Knowledge Visualization to Enhance Human-Agent Situation Awareness Within a Computational Recognition-Primed Decision System. Presented at the 5th IEEE Workshop on Situation Management at MILCOM, Boston, MA, 18–21 October 2009.

6. Torgerson, W. S. Multidimensional Scaling: I. Theory and Method. *Psychometrika* **1952**, *17*, 401–419.

7. Young, F. *Understanding Multidimensional Scaling Kotz-Johnson (Ed.) Encyclopedia of Statistical Sciences*, vol. 5; John Wiley & Sons, Inc.: Hoboken, NJ, 1985.

8. Cox, T.; Cox, M. *Multidimensional Scaling*, 2nd ed.; CRC Press: Boca Raton, FL, 2001.

9. Shepard, R. N. The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function. *Psychometrika* **1962**, *27*, 125–139, 219–246.

10. Kruskal, J. B. Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika* **1964**, *29*, 115–129.

11. Heady, R.; Lucas, J. *PERMAP Operation Manual*; University of Louisiana at Lafayette and Agnes Scott College, GA, March 2007.

12. Stevens, S. S. On the Theory of Scales of Measurement. *Science* **1946**, *103* (2684), 677–680.

13. Gower, J. A. General Coefficient of Similarity and Some of Its Properties. *Biometrics* **1971**, *27* (4), 857–871.

14. Coppock, S.; Mazlack, L. Multi-Modal Data Fusion: A Description. Presented at the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems,Wellington, New Zealand, 20–25 September 2004.

15. Ji, Y.; Massanari, R.; Ager, J.; Yen, J.; Miller, R.; Ying, H. A Fuzzy Logic-Based Computational Recognition-Primed Decision Model. *Information Sciences* **2007**, *177* (20), 4338–4353.

16. Podani, J. Extending Gower's General Coefficient of Similarity to Ordinal Characters. *Taxon.* **1999**, *48*, 331–340.

17. Hanratty, T.; Richardson, J. *A Visual Analytic for High-Dimensional Data Exploitation: The Heterogeneous Data-Reduction Proximity Tool*; ARL-TR-6502; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, August 2013.

18. Heilman, E. *A Military Vignette for a Heterogeneous Data Proximity Tool (HDPT) Study*; ARL-TR-6489; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, June 2013.

19. Wang, W. New Similarity Measures on Fuzzy Sets and on Elements. *Fuzzy Sets Syst*. **1997**, *85* (3), 305–309.

20. Hanratty, T.; Hammell II, R.; Heilman, E.; Dumer, J. Capturing the Value of Information in Complex Military Environments. Presented at the 9th IEEE international Conferernce of Fuzzy Systems, Brisbane, Australia, 10–15 June 2012.

Appendix. HDPT Survey Form

HDPT C4ISR OTM User Survey
C4ISR MOD On-The-Move Exercise E12

Age: ___ Years MOS: _________ Rank: ________ Time in Service: _____ Yrs

What is your current Duty Position? ____

Are you Color Blind? _____ Computer Use Confidence: 1 2 3 4 5 (1-low; 5-high)

Have you been deployed? _____ if *yes*, number of times _______

Have you received military intelligence training? _____
If *yes*, what courses or informal training?

1. Please circle the number below that best describes how ***well*** you were able to use this tool.

1	2	3	4	5

VERY POOR AVERAGE VERY WELL

2. Please circle the number below that best describes your ***awareness*** of the evolving High Value Individual (HVI) intelligence picture.

1	2	3	4	5

NOT AT ALL AWARE SOMEWHAT AWARE VERY AWARE

3. Please circle the number that best describes the HDPT's ***usefulness*** in tactical deployments.

1	2	3	4	5

NOT AT ALL AWARE SOMEWHAT AWARE VERY AWARE

Comments:

List of Symbols, Abbreviations, and Acronyms

Add	address
ARL	U.S. Army Research Laboratory
C4ISR	Command, Control, Communication, Computers, Intelligence, Surveillance, and Reconnaissance
CR	criminal record
DB	database
DCGS-A	Distributed Common Ground System-Army
E12	Event 12
HDPT	heterogeneous data proximity tool
HN	host nation
ID	identification
IED	improvised explosive device
JSON	JavaScript Object Notation
MDS	multidimensional scaling
MOS	military occupational skill
OTM	on the move
OWF	Ozone Widget Framework
REST	Representation State Transfer
SAF	small arms fire
TCP/IP	Transmission Control Protocol/Internet Protocol
VBIED	vehicle borne improvised explosive device

NO. OF COPIES	ORGANIZATION
1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DIRECTOR US ARMY RESEARCH LAB IMAL HRA
1 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	ARMY G2 D WALSH
1 (PDF)	MULTISOURCE INFORMATION FUSION RESEARCH PROFESSOR (EMERITUS) J LLINAS
	ABERDEEN PROVING GROUND
1 (PDF)	US ARMY CERDEC I2WD RDER IWP D PORTER
33 (15 HC, 18 PDF)	DIR USARL RDRL CII A S H YOUNG RDRL CII B B BROOME L TOKARCIK R WINKLER RDRL CII C B BODT E BOWMAN F BRUNDICK J DUMER T HANRATTY E HEILMAN (15 HC, 1 PDF) S KASE M MITTRICK A NEIDERER K OGAARD J RICHARDSON H ROY M THOMAS RDRL CII T V M HOLLAND